텍스쳐 추출시 제한된 수의 참여 영상을 이용한 Multi-view 영상 개선 알고리듬

김도현† · 양영일††

요 약

본 논문에서는 텍스쳐 추출시 제한된 수의 참여 영상을 이용한 multi-view 영상으로부터 가장 좋은 텍스쳐를 추출하는 효과적인 알고리듬을 제안하였다. 기존의 알고리듬이 정규화된 물체 공간에서 X-Y 평면을 삼각패치로 나누고 아핀 변환에 기반한 변이 보상 모델을 이용하여 삼각패치의 텍스쳐를 추출하였다. 본 논문에서는 기존의 방법과 달리 텍스쳐 추출시 참여 영상의 수를 제한하여 multi-view 영상으로부터 가장 좋은 텍스쳐를 추출하였다. *Dragon, santa, city* 그리고 *kid*의 multi-view 영상 세트에 대해 실험한 결과 제안된 알고리듬으로 텍스쳐를 추출한 후 이로부터 복원된 영상의 신호 대 잡음비(SNR)는 기존의 알고리듬으로 처리된 후 복원된 영상의 신호 대 잡음비보다 평균 $0.2dB$ 정도 개선된 결과를 얻을 수 있었다. 제안된 방법으로 부호화된 데이터로부터 복원된 영상은 기존의 방법으로 부호화된 데이터로부터 복원된 영상보다 영상의 화질이 개선됨을 관찰 할 수 있었다.

An Algorithm for the Multi-view Image Improvement with the Restricted Number of Images in Texture Extraction

Do-Hyun Kim† and Yeong-Yil Yang††

ABSTRACT

In this paper, we propose an efficient multi-view image coding algorithm which finds the optimal texture from a restricted number of multi-view image. The X-Y plane of the normalized object space is divided into the triangular patches. The depth of each node is determined by appling a block based disparity compensation method. Thereafter the texture of each patch is extracted by appling an affine transformation based disparity compensation method to the multi-view images. We reduced the number of images needed to determine the texture compared to traditional methods which use all the multi-view image in the texture extraction. The experimental results show that the SNR of images encoded by the proposed algorithm is better than that of images encoded by the traditional method by the approximately $0.2dB$ for the test sets of multi-view image called *dragon, santa, city* and *kid*. Image data recovered after encoding by the proposed method show a better visual results than after using traditional method.

1. 서 론

음성, 영상, 데이터가 하나의 미디어로 처리되고 전송되는 멀티미디어의 시대가 도래하였다. 멀티미디어 관련 기술의 계속적인 발전은 1.2Gbps의 전송을 요구하는 고선명 텔레비젼 방송을 가능하게 하였다. 이런 발전추세를 보면 고선명 텔레비젼보다 더 많은 데이터 량의 전송처리가 요구되는 3차원 텔레비전은 고도의 인간 감각에 좀더 다가갈 수 있는 영상을 보내어 새로운 문화를 창조해 가는 다음 세대의 미디어로 부각될 것이다.

3차원 영상코딩 기술은 스테레오 영상처리[1-3]

본 연구는 정보통신부에서 지원하는 대학기초연구지원사업으로 수행되었슴.

† 경상대학교 전자재료공학과
†† 경상대학교 전기전자공학부 부교수

와 더불어 임의의 위치에서 보았을 때 영상을 재구성하는 것이 용이한 multi-view 영상처리에 관한 연구가 진행되었다[4-7]. Multi-view 영상처리는 여러 장의 영상으로부터 깊이와 텍스쳐로 추출한 후 저장 또는 전송한 후 복원하는 영상 압축 기술이다. Fujii 등은[5] 정규화 좌표계를 사용하여 물체를 표현한 후 블록에 기반한 변이 보상 방법을 사용하여 깊이와 텍스쳐를 동시에 구하거나, 아휜 변환에 기반한 변이 보상 모델을 사용하여 깊이와 텍스쳐를 구하였다. 그리고 김도현 등은[6] 깊이와 텍스쳐를 추출하는 단계를 분리함으로써, 깊이와 텍스쳐 모두 지역적인 값에 의해 결정되고 artifact가 없는 복원영상을 얻는 알고리듬을 제안하였다. 이들 방법에서는 깊이를 구한 후, 주어진 깊이에서 텍스쳐를 추출할 때 해당되는 깊이에 관여한 모든 영상들의 평균값으로부터 결정하였다. 많은 수의 영상들이 텍스쳐 추출에 기여하므로 흐려진 영상을 얻게된다.

본 논문에서는 텍스쳐를 깊이 결정에 관여한 영상 중 일부만을 사용하여 추출하는 알고리듬을 제안한다. 제안된 방법은 첫 째 정규화 공간에서 X-Y 면을 삼각패치로 나눈 후 깊이 추출 단계에서 각 노드를 둘러쌓는 사각형 영역에 대하여 블록에 기반한 변이 보상 방법을 적용함으로서 노드의 깊이를 결정한다. 마지막 단계로 삼각패치의 텍스쳐는 아휜 변환에 기반한 변이 보상 모델을 이용하여 깊이 결정에 관여한 영상 중 우선 순위가 높은 일부 영상으로부터 구한다. 모든 영상으로부터 추출된 텍스쳐로부터 복원된 영상에 비하여 텍스쳐 추출시 참여 영상의 수를 제한하여 얻어진 정보로부터 복원함으로서 화질이 개선된 복원 영상을 얻었다.

제 2 장에서는 multi-view 영상과 multi-view 물체공간의 기하학적 관계를 기술하였고, 제 3 장에서는 본 논문에서 제안한 알고리듬에 대해 기술하였다. 실험 결과를 제 4 장에 설명하였고, 제 5 장에서 결론을 내렸다.

2. Multi-view 영상과 물체공간의 기하학적 관계

여러 대의 카메라를 사용하여 물체를 촬영하거나, 한 대의 카메라를 이동시키면서 물체를 촬영하였을 때, 동일한 물체에 대하여 여러 장의 영상이 얻어진다. 이 때 얻어진 여러 장의 영상을 multi-view 영상 세트라 한다. 그림 1은 *dragon* 물체를 촬영하여 얻어진 multi-view 영상 세트를 보여준다. 카메라는 $+z$ 축을 향하여 $(-2D, 0, z_0)$, $(-D, 0, z_0)$, $(0, 0, z_0)$, $(D, 0, z_0)$과 $(2D, 0, z_0)$에 놓여있다.[8] 뷰포인트에서 $+z$ 방향으로 f만큼 떨어진 곳에 상이 맺히게 되며, 상이 맺히는 x-y 평면을 영상 평면 x_v-y_v라 한다.

그림 1. Multi-view 영상 세트

그림 2는 multi-view 영상의 기하학적 관계를 보여준다. 그림 2에서 왼쪽 그림은 물체공간에서 multi-view 영상을 획득하는 과정이다. 카메라는 $y=0$, $z_0=-F$ 인 위치에 x축으로 D만큼씩 떨어져 놓여 있다. 그러므로 뷰포인트는 $V(nD, 0, -F)$로 표현되고, n은 정수이며 $n=0$ 이면 z 축상의 뷰포인트를 나타낸다. n번째 영상은 뷰포인트 $V(nD, 0, -F)$에서 얻어진 영상을 말한다.
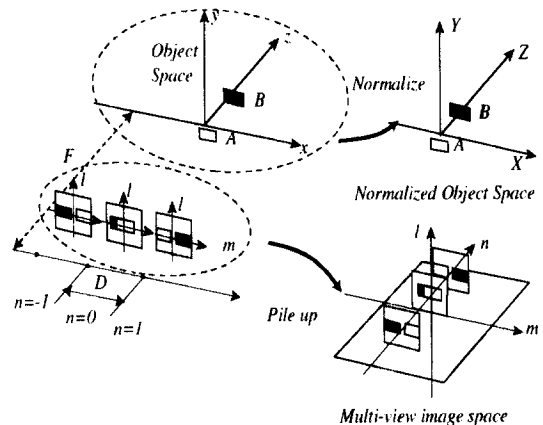
그림 2. 정규화 물체 공간과 multi-view 영상

물체공간 좌표로 표현된 물체가 그림 2의 오른쪽 위 부분에 보인 것처럼 정규화 공간 좌표계로 변환되어 표현되고, 정규화 공간상의 점 (X, Y, Z)는 식 (1)과 식 (2)에 의하여 n번째 영상 $I_n(m, l)$에 투영된다[5,6]. 식(1)은 multi-view 영상간의 좌표를 Z 픽셀(pixel) 만큼의 변이(disparity)로 정규화하는 것이다.

$$m = X + nZ \tag{1}$$

$$l = Y \tag{2}$$

그림 2의 오른쪽 아래 부분에 보인 것처럼, $I_n(m, l)$를 n 방향에 따라 배열하여 형성되는 n-m-l 공간을 multi-view 영상 공간(multi-view image space)이라 한다. Multi-view 영상 공간의 특징을 이용해 multi-view 영상의 압축과 복원을 수행한다.

3. Multi-view 영상에서 제한된 수의 참여 영상을 이용한 텍스쳐 추출 알고리듬

이 장에서는 multi-view 영상의 압축과 복원을 수행하기 위해 본 논문에서 제안하는 깊이와 텍스쳐를 추출하는 방법을 설명한다. 그림 3은 5개의 multi-view 영상으로부터 깊이와 텍스쳐를 추출한 후, 깊이와 추출된 텍스쳐로부터 영상을 복원하는 과정을 보여준다. Multi-view 영상 코딩에서 정규화 공간에서의 물체를.생성할 때 전체 노드의 깊이를 결정하기 위해 노드 주위의 값을 고려한 블록에 기반한 변이 보상 모델로 최적의 깊이를 결정한다[6]. 이때 입력 영상 모두가 깊이 결정에 참여하나 더 좋은 텍스쳐를 추출하기 위해서 참여 영상의 수를 제한하였다. 모든 입력 영상을 이용하여 깊이를 결정하는 과정을 실선으로 표현하였고 참여 영상을 제한하여 텍스쳐를 추출하는 과정을 점선으로 표현하였다. 아래 그림 3에서 5개의 multi-view 영상 (n=-2,-1, 0, 1, 2) 중 3개의 영상으로 텍스쳐 추출을 제한한다면 참여 영상의 수는 3개 이하가 될 수 있다. 참여하는 영상을 선택하는 방법은 영상의 휘도의 분산값을 구한 후 휘도의 분산값이 작은 순서대로 나열한 다음 분산값이 작은 영상들의 휘도값을 평균하여 텍스쳐를 추출한다. 분산값이 큰 영상들은 영상들의 평균 휘도값과 차이가 큰 영상임을 의미하므로 이 영상들이 텍스쳐 추출에 참여하면 흐려진 텍스쳐 영상이 추출된다. 그림 3에서는 multi-view 영상 휘도 분산값의 크기가 n=1, 2, 0, -1, -2 의 순서로 나열되면, 텍스쳐는 분산값이 작은 n= 0, 1, 2 multi-view 영상의 휘도의 평균값이다.

그림 3. 깊이와 텍스쳐 추출을 이용한 multi-view 영상의 압축과 복원

알고리듬 1은 제안한 multi-view 영상 압축 알고리듬의 pseudo code 이다. 그림 4는 정규화 공간과 multi-view 사이의 관계를 나타낸다. 첫 단계는 정규화 공간에서 Z=0인 X-Y 평면을 삼각패치로 나눈다. 둘째 단계는 깊이 추출 단계로 그림 4의 삼각패치의 각 노드의 깊이가 결정된다. 세 번째 단계는 각 삼각패치의 텍스쳐를 구하는 단계이다. 그림 4(a)의 삼각패치의 각 노드 d의 깊이를 결정하기 위해서, 노드 d를 둘러쌓은 사각형 영역 W_d를 Z방향으로 움직인다. 사각형 영역 W_d가 주어진 Z에 놓이면 multi-view 영상에서 이에 대응되는 사각형이 그림 4(a)에서처럼 구해진다. 식 (3)과 식 (4)에 의하여 휘도 분산값 $I_{var}(X, Y, Z)$을 구한다. 가능한 Z값에 대하여 구하여진 휘도 분산값 중 가장 적은 휘도 분산값을 갖는 Z의 위치를 노드의 깊이로 결정한다. 이와같이 블록에 기반한 변이보상 방법을 사용하여 노드의 깊이를 결정한다. 정규화된 공간상의 각 삼각패치의 위치가 결정되었으므로 아휜변환을 이용하면 삼각패치에 대응되는 multi-view 각각의 영상에서의 삼각패치가 구해진다. 정규화 공간상의 삼각패치 P에 대하여 대응되는multi-view 공간상의 삼각패치들을 $p_1, p_2, \cdots p_n$ 이라 하자. 삼각패치 p_i의 텍스쳐를 식 (3)

을 사용하여 평균값 $I_{avg}(X, Y, Z)$를 구한다. 각 삼각패치에 대하여 삼각패치의 휘도값 I_{var}과 평균값 I_{avg}과의 차의 절대값 $Diff(p_i)$를 구한다. 절대값 $Diff(p_i)$가 작은 순서대로 삼각패치를 나열한 후, 정해진 수 만큼의 삼각패치를 선택한다. 선택된 삼각패치에 대하여 식 (3)의 방법으로 평균값을 구한 후 정규화 공간의 삼각패치의 텍스쳐를 배정한다.

$$I_{avg}(X, Y, Z) = \frac{1}{(2N+1)} \sum_{n=-N}^{N} I_n(X+nZ, Y) \quad (3)$$

$$I_{var}(X, Y, Z) = \frac{1}{2N+1} \sum_{n=-N}^{N} (I_n(X+nZ, Y) - I_{avg}(X, Y, Z))^2 \quad (4)$$

알고리듬 1 : Multi-view 영상처리 알고리듬의 슈도 코드

Input : *Multi-view 영상 세트*
Output : *깊이와 텍스쳐*
STEP1 : 정규화 공간에서 Z=0인 X-Y 평면을 삼각패치로 나눈다. (그림 4 참조) 삼각패치 노드의 Z=0 이다.
STEP2 : 각 노드에서 노드를 둘러싼은 사각형 영역(블록)에 대하여 노드의 Z 값을 변화시키면서 블록에 기반한 변이 모델을 이용하여 텍스쳐 분산값이 가장 적은 노드의 깊이, Z값을 찾는다.
STEP3 : Multi-view 영상에서 깊이 결정에 관여한 영상 중에서 일부만 아휜 변환을 이용하여 각 삼각패치에 대응되는 텍스쳐들의 평균값을 구한 후, 삼각패치의 텍스쳐로 할당한다.

4. 실 험 결 과

본 논문에서 제안한 알고리듬을 실험하기 위하여 컴퓨터로 생성한 *dragon* 영상과 일본 Tsukuba 대학에서 제공한 실 영상 *kid*, *city*, *santa* 영상을 사용하였다. 각 영상은 9개의 256×256 그레이 영상으로 이루어져 있다.

본 논문에서는 Fujii의 방법과 9, 5, 3개의 참여 영상으로부터 텍스쳐를 추출하는 실험을 하여 식 (5)를 이용하여 영상의 신호 대 잡음비(SNR)를 비교하였다.

$$SNR = 20\log_{10}\frac{255}{\sqrt{MSE}} \quad (5)$$

여기서 MSE는 복원 영상과 원래 영상의 에러 제곱의 평균을 나타낸다.

그림 5는 *dragon* 영상에 대하여 Fujii 알고리듬과 제안한 알고리듬을 사용하여 압축한 후 복원한 multi-view 영상 중 n=-3, 0, 3 번째 영상을 보여 주고 있다. 그림 5(b)는 Fujii가 제안한 알고리듬에 의해 복원된 영상을 보여 주고 있다. 그리고 그림 5(c), 그림 5(d), 그림 5(e)는 각각 참여 영상이 9개, 5개, 3개로 제한하여 텍스쳐를 추출한 후 복원된 *dragon*

(a) (b)

그림 4. Multi-view 영상처리에서 (a) 깊이와 (b) 텍스쳐 결정

(a)

(b)

(c)

(d)

(e)

그림 5. *Dragon* multi-view 영상 세트 (a) 인코딩하기 전의 원 영상(n=-3, 0, 3) (b) Fujii 방법에 의해 복원된 영상 (c) 9개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상 (d) 5개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상 (e) 3개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상

(a)

(b)

(c)

(d)

(e)

그림 6. *Kid* multi-view 영상 세트 (a) 인코딩하기 전의 원 영상(n=-3, 0, 3) (b) Fujii 방법에 의해 복원된 영상 (c) 9개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상 (d) 5개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상 (e) 3개 영상을 텍스쳐 추출에 사용하였을 때 복원된 영상

영상을 보여준다. 그림 6에서는 각 알고리듬을 사용하여 얻어진 실 영상 *kid*를 보여준다. 표 1은 *dragon* 영상에 대하여 Fujii가 제안한 알고리듬과 텍스쳐 추출시 참여 영상의 수를 제한하여 압축 후 복원하였을 때 신호 대 잡음비를 나타내고, 그림 7에 그래프로 표시하였다. 표 2는 *kid* 영상에 대하여 Fujii가 제안한 알고리듬과 텍스쳐 추출시 참여 영상의 수를 제한하여 압축 후 복원하였을 때 신호 대 잡음비를 나타내고있다. 깊이와 텍스쳐 분리 단계의 알고리듬은 Fujii가 제안한 모델보다 *dragon*과 *kid* 영상에서 평

표 1. 텍스쳐 추출시 제한된 수의 참여 영상에 따른 *dragon* 영상의 신호 대 잡음비 비교

Number of image	SNR(dB)			
	Fujii 알고리듬	제안한 알고리듬		
		9개 영상참여	5개 영상참여	3개 영상참여
-4	25.1	26.5	26.4	26.3
-3	26.4	28.9	28.5	28.2
-2	27.1	30.4	30.3	30.4
-1	29.3	33.8	34.7	34.8
0	30.5	34.7	36.6	37.4
1	30.3	33.6	34.8	34.8
2	28.2	31.4	31.7	30.6
3	26.4	29.5	28.4	28.1
4	24.8	27.2	26.5	26.2

그림 7. 참여 영상의 수에 따른 dragon 영상의 신호 대 잡음비의 비교

균 신호 대 잡음비는 약 $3dB$ 향상됨을 알 수 있다. 제안한 알고리듬에서 참여 영상이 5일 때 가장 좋은 텍스쳐를 얻을 수 있었다. 참여 영상이 5일 때의 신호 대 잡음비는 참여 영상이 9일 경우에 비해 $0.2dB$ 정도 좋은 결과를 얻었다. 텍스쳐 추출시 영상을 제한하여 얻어진 정보로부터 복원된 영상은 모든 영상으로부터 추출된 텍스쳐로부터 복원된 영상보다 영상의 화질이 개선됨을 관찰 할 수 있었다. 그림 8은 참여 영상이 9개와 5개인 경우 복원된 영상에서 *dragon* 영상을 자세히 보여준다. 참여 영상이 5개인 경우 복

표 2. 텍스쳐 추출시 제한된 수의 참여 영상에 따른 *kid* 영상의 신호 대 잡음비 비교

Number of image	SNR(dB)			
	Fujii 알고리듬	제안한 알고리듬		
		9개 영상참여	5개 영상참여	3개 영상참여
-4	23.8	27.4	27.3	27.0
-3	24.9	29.2	29.1	27.8
-2	27.2	30.2	30.4	29.9
-1	28.4	31.7	32.2	32.1
0	29.2	32.3	33.3	33.4
1	27.8	31.8	32.4	32.4
2	26.3	30.4	31.1	30.3
3	24.1	28.5	28.3	28.0
4	23.3	27.6	27.4	27.2

원된 영상의 경계선 등은 뚜렷하나, 참여 영상이 9개인 경우는 *dragon* 물체의 발과 머리 등의 경계선이 흐려짐을 알 수 있다.

5. 결 론

본 논문에서는 텍스쳐 추출시 제한된 수의 참여 영상을 이용한 개선된 multi-view 영상코딩 알고리듬을 제안하였다. 제안한 알고리듬은 Windows NT에서 Visual C++를 사용하여 구현하였고, 9개의 영상, 5개 영상 3개의 영상 참여로 이루어진 multi-view 영상 셋에 대하여 Pentium II를 사용하여 수행하였을 때 인코딩 할 때 약 12~15초 소요되고, 디코딩시 약 2초가 소요되었다. 기존의 알고리듬에서 깊이 정보와 텍스쳐 결정하는데 9개의 영상이 참여한 반면 본 논문에서는 깊이 정보를 결정하기 위해 9개의 영상이 사용되었고 텍스쳐 추출에는 5개의 영상을 사용하여 가장 좋은 신호대잡음비를 얻었다. 제안된 알고리듬에 의해 복원된 영상들은 기존의 알고리듬으로 복원된 영상들에 비해 평균 신호 대 잡음비가 약 0.2 dB 정도 향상되었다. 결론적으로 텍스쳐 추출시 영상을 제한하여 얻어진 정보로부터 복원된 영상은 모든 영상으로부터 추출된 텍스쳐로부터 복원된 영상보다 경계선 등의 영상 화질이 개선됨을 관찰

(a) (b)

그림 8. (a) 텍스쳐 추출 시 9개의 참여 영상 이용하여 복원한 *Dragon* 영상과 (b) 5개의 참여 영상을 이용하여 복원한 *Dragon* 영상

할 수 있었다.

참 고 문 헌

[1] U. R. Dhond and J. K. Aggarrwal, "Structure from Stereo - A Review," *IEEE Trans. Systems, Man, Cybernetics*, Vol. 19, No. 6, pp. 1489-1510, 1989.

[2] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-2, pp.333-340, July 1980.

[3] Y. Ohta and T. Kanade, "Stereo by intra-and inter-scanline search using dynamic programing," *IEEE Trans. PAMI*, Vol. 7, No. 2, pp.139-154, 1985.

[4] M. E. Lukas, "Predictive Cording of Multi-Viewpoint Image Sets," *IEEE ICASSP'86*, pp.521-524, 1986.

[5] T. Fujii and H. Harashima, "Data Compression and Interpolation of Multi-View Image Set," *IEICE Trans. Inf. & Syst.*, E77-D, 9, pp.987-995, 1994.

[6] 김도현, 최동준, 양영일, "Multi-view 영상코딩에서 영상 개선 알고리듬," 대한전자공학회논문지, 제35권, S편 제7호, pp.53-61, 1998.

[7] H. Aydinoglu and M. H. Hayes III, "Compression of Multi-View Images," *IEEE ICIP'94*, pp. 385-389, 1994.

[8] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-Plane Image Analysis : An Approach to Determining Structure from Motion," *International Journal of Computer Vision*, 1, pp. 7-55, 1987.

김 도 현

1970년 4월 11일생
1996년 2월 경상대학교 전자재료공학과 졸업(학사)
1998년 2월 경상대학교 전자재료공학과 졸업(석사)
1998년~1999년 경상대학교 항공기부품기술연구센터 연구원
1999년~현재 경상대학교 전자재료공학과 박사과정
관심분야 : 3차원 영상처리신호와 ASIC 설계임

양 영 일

1983년 2월 경북대학교 전자공학과 졸업(학사)
1985년 한국과학기술원 전기 및 전자공학과 졸업(석사)
1989년 한국과학기술원 전기 및 전자공학과 졸업(박사)
1990년~현재 경상대학교 전기전자공학부 부교수
1994년 1월~1995년 1월 UC, Irvine 교환교수
관심분야 : VLSI&CAD, 영상신호처리 등